Re: Distributing n points in space

Source: http://sci.tech−archive.net/Archive/sci.math/2008−06/msg01036.html

- *From*: Ray Koopman <koopman@xxxxxx>
- *Date*: Wed, 11 Jun 2008 17:03:28 −0700 (PDT)

On Jun 11, 4:02 pm, Robert Israel
<isr...@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote:

> I have n points i don't have any coordinates, only thing i have is distance matrix por each pair of points distance from one to the other, i want to put them all in k<n dimensional space, it is probably impossible to do it precisly what i want is to distribiute points in space in in the way that minimize sum of diference between proper distance and the real one. Is there any solution of this, does the problem have any official name?

Consider the Gram matrix G:
G_{ij} = (x_i).(x_j)
G must be positive semidefinite. The n x n positive semidefinite matrix G is the Gram matrix for n points of R^k if and only if G has rank <= k, and these points can be obtained using the diagonalization of G. If G has some negative eigenvalues or rank > k, the best approximation, in some sense, is obtained by using the (up to) k largest positive eigenvalues and eigenvectors.

Now given the distance matrix D:
D_{ij} = |x_i − x_j|
we can designate one of the points (say x_n) as the origin, and then
G_{ij} = (D_{in}^2 + D_{jn}^2 − D_{ij}^2)/2

For example, suppose you have the 4 x 4 distance matrix
[0 2 2 3]
[2 0 3 2]
D = [2 3 0 1]
[3 2 1 0]

and you want k=2. This leads to

[9 4.5 3 0]
[4.5 4 −2 0]
G = [3 −2 1 0]
[0 0 0 0]

which has a negative eigenvalue. If lambda_1 and lambda_2 are the two
positive eigenvalues and u_1 and u_2 the corresponding orthonormal
eigenvectors, then we take the points of R^2 corresponding to the
rows of the matrix with columns sqrt(lambda_1) u_1 and sqrt(lambda_2) u_2,
namely

[−3.01692215332068070, .520107363143339496],
[−1.58230163082378516, −1.45887821879530888],
[−.540150726393245684, 1.36862161370595770],
[0., 0.]

which would have the distance matrix
[0. 2.444283121 2.618085777 3.061426293]
[2.444283121 0. 3.013442187 2.152209123]
[2.618085777 3.013442187 0. 1.471355813]
[3.061426293 2.152209123 1.471355813 0.]
−−

Robert Israel isr...@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
Department of Mathematics http://www.math.ubc.ca/~israel
University of British Columbia Vancouver, BC, Canada

The fit will sometimes be slightly better if you take the origin at
the centroid. To do this, let G be −.5 times the double−centered
matrix of squared distances; i.e.,
G_{i,j} = −.5 * (DD_{i,j} − DD_{i,.} − DD_{.,j} + DD_{.,.}),
where DD contains the squared distances,
and a "." denotes averaging over the corresponding index.

For the example data, the recovered distances in 2 dimensions are

0. 2.22918 2.30898 3.00478
2.22918 0. 3.00478 2.30898
2.30898 3.00478 0. 1.65858
3.00478 2.30898 1.65858 0.
.